

Statistical Dynamics of Regional Populations and Economies

Jie Huo^{1,2}, Xu-Ming Wang^{1,2,*}, Rui Hao^{1,2}, and Peng Wang¹

¹School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan, 750021, P. R. China

²Ningxia Key Laboratory of Intelligent Sensing for Desert Information, Yinchuan, 750021, P. R. China

*wang_xm@126.com (Xu-Ming Wang)

ABSTRACT

Quantitative analysis of human behavior and social development is becoming a hot spot of some interdisciplinary studies. A statistical analysis on the population and GDP of 150 cities in China from 1990 to 2013 is conducted. The result presents the cumulative probability distribution of the populations and that of the GDPs obeying the shifted power law, respectively. In order to understand these characteristics, a generalized Langevin equation describing variation of population is proposed based on the correlation between population and GDP as well as the random fluctuations of the related factors. The equation is transformed into the Fokker-Plank equation to express the evolution of population distribution. The general solution demonstrates a transform of the distribution from the normal Gaussian distribution to a shifted power law. It in fact suggests a critical point of time at which the transform takes place. The shifted power law distribution in the supercritical situation is qualitatively in accordance with the practical result. The distribution of the GDPs is derived from the well-known Cobb-Douglas production function. The result presents a change, in supercritical situation, from a shifted power law to the Gaussian distribution. This is a surprising result—the regional GDP distribution of our world will be the Gaussian distribution some day in the future. The discussions based on the changing trend of the economy growth suggest it will be true. Therefore, these theoretical attempts may draw a historical picture of our society in the aspects of population and economy.

Introduction

In the past decade, migration behavior of population has drawn great interest of scientists. Many physicists involved in the studies to discover the characteristics and corresponding mechanism. They have explored the laws that dominate the relationship between the migrated population and the related factors, and studied the distributions of the related aspects such as migrated population, migrating distances and per economic incomes, etc. The groundbreaking field work based on practical observations was contributed by E. G. Ravenstein in 1885; and the well-known 7-rules that govern population migration has been proposed¹.

From then, scientists have investigated this issue from two different points of view. One focuses on finding out the driving factor of the migrations based on the knowledge of economics, physiology and sociology; the other emphasizes extracting regularities of migration from observed data and revealing their dynamic origin. The former is represented by the following investigations. The aforementioned factors, associating in some way, are perceived as contributing to migration. Some combinations play the part of repelling force and some others serve as appealing force acting on the immigrating persons². The so-called push-pull theory was proposed based on the motivation of migration, improving the quality of living³. The clearest description of this theory was suggested by Lee via taking the driving factors, amount of persons and directions, into account⁴. A similar contribution was made by Schultz who regarded the migration as a decision-making behavior based on the income-output⁵. The latter is denoted by investigations to reveal the physical essence of migrations. One typical representative is the well-known gravity model, which is in a form analogous to that of the gravity law⁶. It states that the number of people that migrate from one region to another is usually proportional to the product of the numbers of people living in this two regions, and inversely proportional to the distance between the two regions. This work is actually the field-broken one of the dynamic investigations. Following this idea, the intervening opportunities model was proposed to reveal how the number of visitors to a given distance relates to the number of opportunities at that distance and to the number of intervening opportunities⁷. The probabilistic model for urban travel was developed to emphasize the influence of heterogeneity of individuals on the choice of destinations⁸.

Gradually, many studies turn to focus on dynamic details of human movements. The most remarkable one is that practical investigation on the traveling traces via tracking bank notes and the corresponding theoretical interpretation, conducted by Brockman in 2006⁹. It induced a new climax in exploring the spatial characteristics of human movements and the mechanics behind¹⁰. With the aid of the modern information technology, the researchers can precisely capture human motions. For

instances, the data obtained by tracking cellphone calls and that by tracking short-message services has proved that the movements of the cellphone users can be characterized by a Truncated Power Law (TPL) distribution of the trip distances^{10–16}; the GPS data of taxis serves showed that the movements of the passengers can be described by an exponential distribution of the travel distances^{17–23}.

There have been many theoretical attempts to interpret the essential origins of the aforementioned regularities. As far as methodology is concerned, two typical methods are worth mentioning here. The first is dynamical simulation based on constructing the rules to mimic the mobility of human. For instances, the so-called continuous-time random walk model was suggested to reproduce the movements of the bank notes, and drew a conclusion that the motion of the bank note holder can be regarded as a Lévy flight⁹; the preferential return (exploring a new place with a probability and returning to a preferentially visited location with the complementary probability) model was proposed to understand the individual mobility patterns uncovered by cell phone calls¹², and the results match well the individual's daily activity; a similar model, the regular mobility model, was developed to mimic the periodicity and regularity of human daily commuting mode, "home-work-leisure-home"²⁴, and the simulations are in good accordance with those practical observations presented in Refs. 10 and 25; the dynamical model for population migration, driven by economic income and hindered by the cost relating to displace and employment opportunity, can capture the main empirical features, mentioned above, such as the distributions of displaces and population migration sizes, etc.²⁶. The second theoretical method is based on some related physical abstract or reduction of human mobility, and/or utilizing some fundamental principle of Physics to characterize human's behavior referring to the knowledge of the corresponding physical system. The following efforts may be most typical. In order to understand statistical characteristics of population, the idea of the Maxwell-Boltzmann statistics was adopted to interpret the TPL distribution of travel distances. This consideration may be due to the fact that there is similarity between the movement of persons in a city and that of particles in a thermodynamic system. The most probable distribution of equilibrium state can be obtained based on the principle of maximum entropy with the gross cost constraint and the assumptions of the nearly independence and the identity of particles²⁷. To reveal the mechanism dominating scaling properties of human and animal mobility, and to explain what causes the value difference of the scaling exponent between human and animal, the home-return Lévy flight model optimized by information entropy has been built. The exponents can be obtained via optimizing under the constraint of maximizing the information entropy, and indicate that the difference depends on whether there is a home to return, yes for human, not for animal²⁸. The process that some persons leaving home-town and looking for better jobs incarnated by high income, short working time, good working condition, etc., can be regarded as a stochastic process just as the radiation of some particles that are emitting from an origin location, diffusing to look for some locations to release or transport the carried energies or masses. The radiation model catches such a similarity of the physical essence in these two processes. And it works well to predict the population migration from one location to another²⁹. From the discussions above, it can be clear seen if an investigation can grasp analogies between the behaviors of individual/social system and that of particle/physical system, the essential description of human activity may be obtained by translating the corresponding theory or principle of physics. Here we also emphasize another aspect apparently existed in the aforementioned investigations, that is, the scaling distribution feature of human motions may be jointly determined by the stochastic and deterministic driving forces. In fact, similar mechanisms exist widely, in nature, to dominate evolution of systems^{30,31}.

Therefore, in this article, a generalized Langevin equation is constructed to describe the process of population variation including the natural increase and migration; and it can be further transformed into the Fokker-Plank (F-P) equation to investigate the statistical characteristics of population migration. As is well known, the F-P equation is a drift-diffusion equation describing the diffusion of particles in an external potential field, and has been widely extended to investigate many other processes in which the movements of the objects demonstrate the similarity to the random diffusion (Markovian or non-Markovian) of a particle (see the typical representatives in Refs. 32–34). Here, the solution of the F-P equation presents an evolutionary picture of human societies in the aspects of population and GDP. The distribution of the regional populations originally obey the Gaussian-like distribution caused by the random migration. Subsequently, this distribution gradually transformed into a Shifted Power Law (SPL). Based on the Cobb-Douglas (C-D) production function, we suggest a relation function between population size and GDP, and derive the distribution of the regional GDPs. The distribution shows a reverse change from a SPL distribution into the Gaussian distribution that may be regarded as the prediction about the future of our society. If so, what does the Gaussian distribution of the regional GDPs imply? What causes the Gaussian distribution to occur? The following sections will present the detailed discussions.

Results

Statistical characteristics of the regional population sizes and GDPs

Before presenting the theoretical discussions based on the similarity between human and natural substance, we perform a practical statistics to explore the characteristics of population distribution and the related economic regularity. We have collected some basic statistics including regional population size L and GDP Y , of about 150 cities in China. We choose these cities for

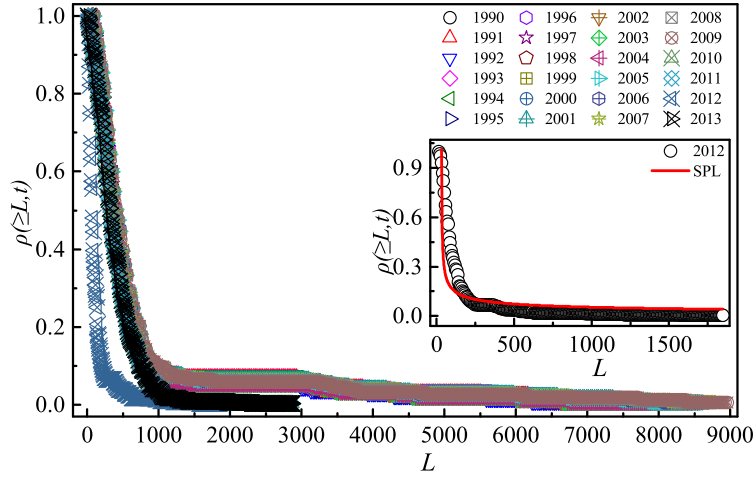


Figure 1. Cumulative probability distribution of population sizes obtain from the China Statistical Yearbook from 1990 to 2013. The inset presents the fitting by SPL relation for the distribution in 2012.

the relative complete data recorded in the China Statistical Yearbook from 1990 to 2013 (<http://www.stats.gov.cn>). Fig. 1 presents the cumulative probability distribution of population sizes at each given time t (in one year) from 1990 to 2013, respectively. The inset re-displays the relation in 2012 chosen as one representative and fitted by the SPL

$$\rho(\geq L) \propto (L + a_1)^{-b_1}. \quad (1)$$

Where $a_1 = -31.673 \pm 0.344$, $b_1 = 0.425 \pm 0.018$. Fig. 2 shows the cumulative probability distribution of GDPs in the same period. The representative relation in 1990 is also well fitted by the SPL and redisplayed by the inset, that is, the distribution function takes a similar form of Eq. 1,

$$\rho(\geq Y) \propto (Y + a_2)^{-b_2}. \quad (2)$$

where $a_2 = -18.141 \pm 0.296$, $b_2 = 0.406 \pm 0.0125$.

It is necessary to point out that these fitting results, SPL distributions, indicate that the evolution of the social system may be jointly driven by deterministic and stochastic dynamic factors according to the statement appeared in Refs. 25 and 29. This conclusion is also in accord with our previous discussions in natural system³⁰ and social system²⁶. The understandings in our theoretical framework will present in the following sections.

Stochastic Dynamical Equation for Evolution of Regional Population

It is well known, the change of regional population is nothing more than the natural increase and migration, and the diffusion of population between regions owes much to the migration. The migration is often induced by higher income at the destination. Therefore, the regional population strongly depends on the regional economy. The economic factors, such as fixed capital investment, flowing capital investment, labor input, marginal propensity to save, potential capacity, etc., vary randomly. Therefore, the evolution of a regional population, L , is a typical stochastic process, and we suggest a generalized Langevin equation to describe this process. It reads in a general form,

$$\frac{dL}{dt} = f(L, t) - g(L, t)\sigma(t) \quad (3)$$

where the first term on the right hand is deterministic corresponding to driving force, and the second stochastic corresponding to the Langevin force induced by the fluctuations of the related factors. Variable $\sigma(t)$ denotes the Gaussian white noise that is characterized by mean value $\langle \sigma(t) \rangle = 0$ and δ -correlation $\langle \sigma(t)\sigma(t') \rangle = 2D\delta(t - t')$.

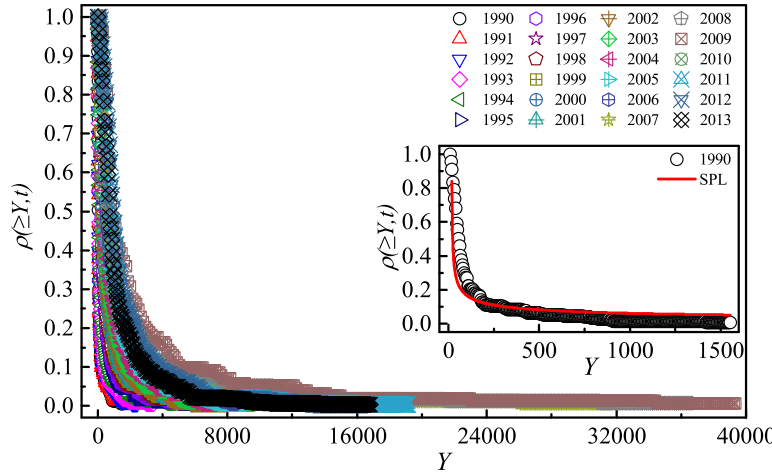


Figure 2. Cumulative probability distribution of GDPs obtained from the China Statistical Yearbook from 1990 to 2013. The inset presents the fitting by SPL relation for the distribution in 1990.

The apparent form of Eq. 1 can be obtained by correlating the well-known production function³⁵ with the economic growth model^{36,37}. The former stresses the dependence of the regional GDP, Y , on the determining factors such as labor input L' , fixed capital investment K and comprehensive technology factor A (which is related to production efficiency), that is,

$$Y = AL'^{\alpha}K^{\beta}. \quad (4)$$

Where α, β is the elasticity coefficient for the labor output and the capital output, respectively. Here we suppose that the labor input is simply proportional to the population, so we have

$$L' = \eta L, 0 < \eta < 1. \quad (5)$$

The latter, the economic growth model, describes economic growth based on a pure quantitative relation linking output and input, which emphasizes that the time variation rate of GDP increases exponentially with time, and is related to the so-called marginal propensity to save and potential social average productivity of investment ξ in the following form

$$\frac{dY}{dt} = \frac{dI}{dt} \frac{1}{\lambda} = (\xi/\lambda)I_0 e^{\xi\lambda t}. \quad (6)$$

Where I, I_0 denotes the capital investment rate and its initial value, respectively. Obviously, the increment of GDP within time interval $t \rightarrow t + \Delta t$ is $A\eta^{-\alpha}(L + \Delta L)^{\alpha}K^{\beta} - A\eta^{-\alpha}L^{\alpha}K^{\beta}$ obtained from Eqs.3 and 4, and $(\xi/\lambda)I_0 e^{\xi\lambda t}\Delta t$ obtained from Eq.6. As a result, we have an equation

$$(\xi/\lambda)I_0 e^{\xi\lambda t}\Delta t = A\eta^{-\alpha}(L + \Delta L)^{\alpha}K^{\beta} - A\eta^{-\alpha}L^{\alpha}K^{\beta}. \quad (7)$$

As $\Delta L(\Delta t)$ has a very small value, the first term on the right hand of the equation can be expanded in Taylor series, and changed into the following form by remaining the linear term

$$(\xi/\lambda)I_0 e^{\xi\lambda t}\Delta t = A\alpha\eta^{-\alpha}L^{\alpha-1}K^{\beta}\Delta L. \quad (8)$$

As $\Delta t \rightarrow 0$ and therefore $\Delta L \rightarrow 0$, the equation can be transformed into the one

$$\frac{dL}{dt} = \frac{\xi I_0 e^{\xi \lambda t} \eta^\alpha}{\lambda \alpha A K^\beta} L^\gamma \quad (9)$$

by setting $\gamma = \alpha - 1$. The right side of the equation is actually the deterministic term of Langevin equation, $f(L, t)$, in which each of the coefficients and variables can be taken as its mean value. Of course, each of the coefficients and variables fluctuates, so the instantaneous form of Eq.9 can be written as

$$\frac{dL}{dt} = \frac{(\xi \pm \Delta\xi)(I_0 \pm \Delta I_0)(\eta \pm \Delta\eta)^{(\alpha \pm \Delta\alpha)} e^{(\xi \pm \Delta\xi)(\lambda \pm \Delta\lambda)t}}{(\lambda \pm \Delta\lambda)(\alpha \pm \Delta\alpha)(A \pm \Delta A)(K \pm \Delta K)^{(\beta \pm \Delta\beta)}} L^\gamma, \quad (10)$$

where $\Delta\xi, \Delta I_0, \Delta\eta, \Delta\alpha, \Delta\lambda, \Delta A, \Delta K, \Delta\beta$ denotes the amplitude of the fluctuation of $\xi, I_0, \eta, \alpha, \lambda, A, K, \beta$, respectively. Truncating appropriately the Taylor series expansion for each term, the equation changes to the following one,

$$\frac{dL}{dt} = \frac{\xi I_0 \eta^\alpha e^{\xi \lambda t}}{\lambda \alpha A K^\beta} L^\gamma \left(1 \pm \left(\frac{\Delta\xi}{\xi} + \frac{I_0}{I_0} + \alpha \frac{\Delta\eta}{\eta} + \Delta\alpha + \Delta\xi + \Delta\lambda - \frac{\Delta\lambda}{\lambda} - \frac{\Delta\alpha}{\alpha} - \frac{\Delta A}{A} - \beta \frac{\Delta K}{K} - \Delta\beta \right) \right). \quad (11)$$

By setting $\phi = \frac{\xi I_0 \eta^\alpha e^{\xi \lambda t}}{\lambda \alpha A K^\beta}$ and $\varepsilon = \frac{\Delta\xi}{\xi} + \frac{I_0}{I_0} + \alpha \frac{\Delta\eta}{\eta} + \Delta\alpha + \Delta\xi + \Delta\lambda - \frac{\Delta\lambda}{\lambda} - \frac{\Delta\alpha}{\alpha} - \frac{\Delta A}{A} - \beta \frac{\Delta K}{K} - \Delta\beta$, the equation is rewritten as

$$\frac{dL}{dt} = \phi L^\gamma \pm \phi \varepsilon L^\gamma, \quad (12)$$

here $\pm \varepsilon$ denotes pure noise caused by the relative fluctuations, and the noise is in terms of $\pm \phi \varepsilon$. Comparing Eq.3 with Eq.12, one can easily obtain the two terms,

$$f(L, t) = \phi L^\gamma, g(L, t) = L^\gamma. \quad (13)$$

Thus, a generalized Langevin equation describing the random growth of regional population is established. It is very clear that such a stochastic equation cannot predict the migration precisely, but can only describe the migration by probability. Therefore, a transform, from the generalized Langevin equation to the F-P equation, is necessary to carry out.

Distribution of Regional Population Sizes

The transform can be conducted as follows. Firstly, one can obtain $L(t + \tau) - L(t)$ via integrating Eq.3 within a small time interval τ (in $t-t + \tau$) along the evolutionary path or integral path. Secondly, one can get n-order moment, for any possible n, as functions $f(L, t)$ and $g(L, t)$ in the expression of $L(t + \tau) - L(t)$ are expanded in series in the neighborhood of $L(t)$, that is,

$$\begin{cases} \langle L(t + \tau) - L(t) \rangle = \{f(L, t) + Dg'(L, t)g(L, t)\}\tau + O(\tau^2), \\ \langle [L(t + \tau) - L(t)]^2 \rangle = 2Dg^2(L, t)\tau + O(\tau^2), \\ \langle [L(t + \tau) - L(t)]^n \rangle = 0 \quad n \geq 3. \end{cases} \quad (14)$$

Where $g'(L, t)$ is the derivative of $g(L, t)$ with respect to L . Finally, the F-P equation can be derived from the moments above via substituting them into the so-called Kramers-Moyal expansion holding over the linear terms. And then we have equation ,

$$\frac{\partial}{\partial t} \rho(L_0|L; t) = -\frac{\partial}{\partial L} \{f(L, t) + Dg'(L, t)g(L, t)\} \rho(L_0|L; t) + D \frac{\partial^2}{\partial L^2} \{g^2(L, t) \rho(L_0|L; t)\} \quad (15)$$

to describe the time evolution of the probability distribution of population sizes. Where $\rho(L_0|L; t)$ denotes the transition probability to measure the probability of a region with population L_0 at $t = 0$ becoming the one with population L after a while t .

Now let's find a general solution of Eq.15, $\rho(L', t|L, t + \tau)$, which denotes the transition probability of the system from time t with population L' to time $t + \tau$ with population L in a small time interval τ . It can be obtained by integrating over τ ,

$$\rho(L', t|L, t + \tau) = [1 - \frac{\partial}{\partial L}(\phi L^\gamma + \gamma D L^{2\gamma-1})\tau + D \frac{\partial^2}{\partial L^2} L^{2\gamma} \tau + O(\tau^2)] \delta(L - L'). \quad (16)$$

The terms of order being equal and higher than 2 may be ignored due to the very small values. So, Eq.16 is simplified as

$$\rho(L', t|L, t + \tau) \approx [1 - \frac{\partial}{\partial L}(\phi L^\gamma + \gamma D L^{2\gamma-1})\tau + D \frac{\partial^2}{\partial L^2} L^{2\gamma} \tau] \delta(L - L'). \quad (17)$$

One may note that the polynomial in the bracket can be regarded as the linear function of time interval τ . It can also be treated as a Taylor series expansion of an exponential function, of which the terms of order 0 and 1 are retained. Therefore, the transition probability can be approximatively "returned" the corresponding exponential function, that is,

$$\rho(L', t|L, t + \tau) \approx \exp[-\frac{\partial}{\partial L}(\phi L^\gamma + \gamma D L^{2\gamma-1})\tau + D \frac{\partial^2}{\partial L^2} L^{2\gamma} \tau] \delta(L - L'). \quad (18)$$

Due to the fact that the probability described by the exponential function is not normalized, the approximately equal sign in the above formula can be changed into an equal sign. After that, introducing the Fourier integral to express the δ -function, one obtains

$$\rho(L', t|L, t + \tau) = \exp[-\frac{\partial}{\partial L}(\phi L^\gamma + \gamma D L^{2\gamma-1})\tau + D \frac{\partial^2}{\partial L^2} L^{2\gamma} \tau] \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu(L-L')} du. \quad (19)$$

According to the differential relation of the Fourier transform, $\frac{d^n F(\omega)}{d\omega^n} = (-1)^n [t^n f(t)]$, the equation can be written in the form

$$\begin{aligned} \rho(L', t|L, t + \tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-iu(\phi L^\gamma + \gamma D L^{2\gamma-1})\tau - u^2 D L^{2\gamma} \tau + iu(L - L')] du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-D L^{2\gamma} \tau (u^2 - iu \frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]}{D L^{2\gamma} \tau} \\ &\quad + \frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]^2}{4(D L^{2\gamma} \tau)^2}) - \frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]^2}{4D L^{2\gamma} \tau}] du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-D L^{2\gamma} \tau (u - i \frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]}{2D L^{2\gamma} \tau})^2 - \frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]^2}{4D L^{2\gamma} \tau}] du. \end{aligned} \quad (20)$$

Using the Gaussian integral, $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$, one therefore has

$$\begin{aligned} \rho(L', t|L, t + \tau) &= \frac{1}{2\sqrt{\pi D L^{2\gamma} \tau}} \exp(-\frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]^2}{4D L^{2\gamma} \tau}) \\ &= \frac{L^{-\gamma}}{2\sqrt{\pi D \tau}} \exp(-\frac{[(L - L') - (\phi L^\gamma + \gamma D L^{2\gamma-1})\tau]^2}{4D L^{2\gamma} \tau}). \end{aligned} \quad (21)$$

By setting $L' = 0, t = 0$ (the starting point of the evolution is assumed to extrapolate backward to the "primitive society"), and $\tau \rightarrow t$ (extending the small time situation to an arbitrary time situation), so transition probability $\rho(L', t|L, t + \tau)$ can be simplified as $\rho(L, t)$, and then its expression is changed to

$$\rho(L, t) = \frac{L^{-\gamma}}{2\sqrt{\pi D t}} \exp(-\phi \gamma L^{\gamma-1} t + D \gamma (2\gamma - 1) L^{2\gamma-2} t - \frac{[L^{1-\gamma} - \phi t + \gamma D L^{\gamma-1} t]^2}{4D t}). \quad (22)$$

In the following discussions, the values of main parameters are set as $\phi = 0.285, \gamma = 1.47, D = 0.0005$. Obviously, this solution is characterized by a product of two functions. One is in power form and the other in exponent form. This distribution

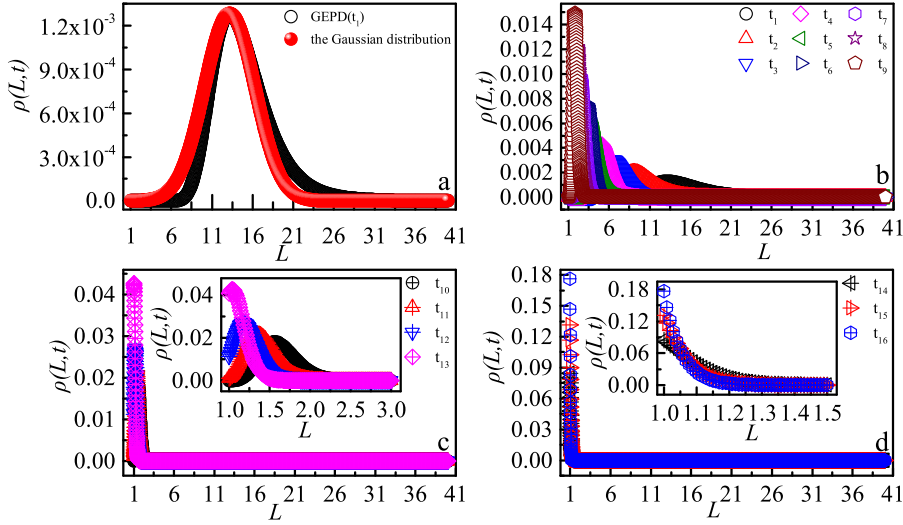


Figure 3. Distribution of the regional populations varies with time. (a) presents the distribution at $t = 1$, the distribution is close to the normal Gaussian distribution; (b) presents the distribution from $t = 1$ to $t = 9$; (c) presents the distribution from $t = 10$ to $t = 13$, its enlargement is shown by the inset; (d) presents the distribution from $t = 14$ to $t = 16$, their enlargement is also shown by the inset.

is called Generalized Exponential Power Distribution (GEPD)^{26,30,38–41}. Similar distribution characteristics can be observed in many real systems, such as temporal and spatial activities of human, biological system, river networks and so on⁴¹. As discussed in Ref. 30, such a distribution law has two equivalent mathematical formulations. One is known as the SPL, expressed as $\rho(x) \propto (x+a)^{-b}$, the other is described as the TPL that is manifestation of $\rho(x) = \mu(x^{\mu-1}/x_0^{\mu})\exp(-(x/x_0)^{\mu})$. The characteristic parameter for the former and the latter is a and μ , respectively. When $a = 0$ ($\mu \rightarrow 0$), the SPL(TPL) relation tends to be a typical power-law relation, whereas when $a \rightarrow \infty$ ($\mu = 1$) the SPL(TPL) relation reduces to a typical exponential distribution.

Here Fig. 3 depicts the final analytical result presented by Eq. 22, the probability distribution of regional population sizes varies with time. This result may reproduce some historical profiles of the human society, that is, exhibit the regional population evolution in the process of the natural increase and migration, as well as the related economy development. Fig. 3(a) shows a typical random distribution similar to the Gaussian distribution (the comparison between the distribution and the standard Gaussian distribution is shown in the Figure), which indicates that human distributes evenly and the "society" runs without structure in the early history. This may be attributed to the low natural population increase rate and the rare and blind migration. Fig. 3(b) demonstrates a gradual change of the distribution with time. The distribution curve loses its early symmetry with respect to the peak point. And the peak point moves to the left. Fig. 3(c) shows the distribution reaches a critical point at $t = 13$. The supercritical distribution completely transforms into the one monotonously descending with population size L . Fig. 3(d) shows the details. Then a question may naturally occur, i.e., what kind is this monotonously descending distribution? Before answering the question, we should check the validation of our theoretical framework via comparing the theoretical result with that obtained practically. So, one has to perform an integration for Eq.22 from L to infinity to yield the cumulative probability distribution of population sizes

$$\rho(\geq L, t) = \int_L^{\infty} \rho(L, t) dL. \quad (23)$$

Fig. 4 presents the cumulative probability in the supercritical situation at $t = 14$. The fitting red solid curve suggests a perfect SPL

$$\rho(\geq L, t) \propto (x + a_3)^{-b_3} \quad (24)$$

with $a_3 = 0.124 \pm 0.00039$, $b_3 = 6.31 \pm 0.0091$. It is clear that the theoretical result is in qualitative accordance with the observed result and that the answer is revealed. It also implies that our theoretical framework can describe the characteristics

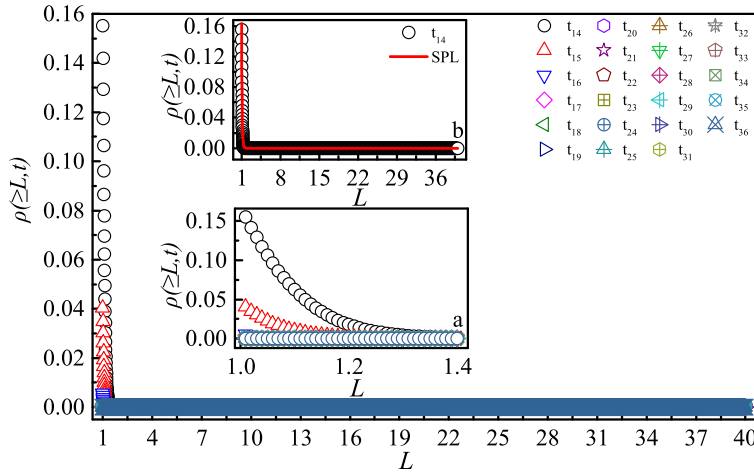


Figure 4. Cumulative probability distribution of populations. The inset (a) is the partial enlargement and the inset (b) the fitting of that at $t = 14$, obeying the SPL.

of such spatial behavior of human. It seems now we can sketch out a dynamical profile of human population distribution evolving with time. In the early period of the evolutionary process, supposing it is the "primitive society", the population migrates or proliferates in a disorderly manner, and therefore leads to the Gaussian distribution. As time goes on, the cluster effect in some regions begins to emerge, so the even "society" signalled by the Gaussian distribution is destroyed, and is gradually transforming to the SPL distribution. This also marks that the society is evolving from disorder to order. Finally, the distribution turns into the SPL, which means that the society has evolved into an order state with a structure of highly inhomogeneous regional populations. In other words, the SPL dominating the population distribution is a sign of a modern society with a developed structure.

Distribution of Regional GDPs

It is generally accepted that the evolution of the regional populations is mainly driven by the benefit pursuing of human to improve the living conditions. So the population closely relates to the economic development level represented by GDP. It especially is so for the modern society. The widely approved relation is known as the C-D production function that can be approximately written as $Y \propto L^\alpha$. One can suggest a relation function for the population and GDP using a δ -function,

$$r(L, Y) = \delta(L - Y^{-\alpha}). \quad (25)$$

With which one can obtain the distribution of the regional GDPs via performing the following integration

$$\begin{aligned} \rho(Y, t) &= \int \rho(L, t) r(L, Y) dL \\ &= \frac{Y^{\alpha\gamma}}{2\sqrt{\pi Dt}} \exp(-\varphi\gamma Y^{-\alpha(\gamma-1)}t + D\gamma(2\gamma-1)Y^{-\alpha(2\gamma-2)}t - \frac{[Y^{-\alpha(1-\gamma)} - \varphi t + \gamma DY^{-\alpha(\gamma-1)}t]^2}{4Dt}). \end{aligned} \quad (26)$$

It is necessary to point out that the relation function between the regional population and GDP relates only to those in supercritical situation. Equally, the integral, distribution of the regional GDPs, is applied only to the supercritical situation. One may say that if the distribution of the regional GDPs can describe the current state, it also can predict the future of the world. The calculating result of Eq. 25 is shown by Fig. 5 as $\alpha = 1.2$ and the other parameters are the same as those in calculating the distribution of population sizes. Fig. 5 (a) shows the distribution of the regional GDPs at the critical point, $t = 13$, which can be perfectly fitted by a SPL. However, it is amazing that the SPL distribution here does not like the distribution of population, staying the way. As is shown by Figs. 5(b) and (c), the distribution is destroyed, and a peak appears on the curve as it is in the supercritical situation. As time goes on, the distribution turns into the Gaussian distribution, Fig. 5(d) proves that the distribution of the regional GDPs is a perfect Gaussian distribution. Furthermore, the peak of the Gaussian distribution moves

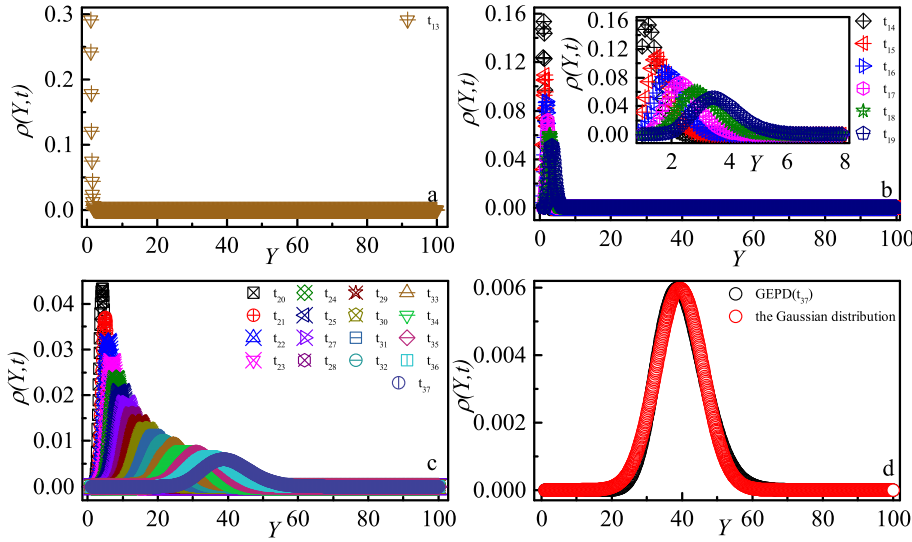


Figure 5. Distribution of the regional GDPs varies with time. (a) presents the distribution at critical point $t = 13$, the distribution is close to the SPL; (b) presents the distribution from $t = 14$ to $t = 19$, the distribution turns into the normal Gaussian distribution at $t = 19$; (c) presents the distribution from $t = 20$ to $t = 37$; (d) re-presents the distribution at $t = 37$, the comparison of it with the normal Gaussian distribution indicates that the distribution keeps well in the Gaussian fashion.

to the right, and drops and broadens meanwhile. The movement to the right implies the increase of the GDP as a whole, the dropping and broadening variation of the peak indicates the homogenization of the distribution. This varying trend may accord with the fact that almost everyone and every region is inclined to seek wealth in the developing process of human society. Perhaps, the Gaussian distribution may be the economic status of our world in future.

It may be important if one can know the cause why the change from the SPL distribution to the Gaussian distribution can occur. However, it is very difficult to provide a physical answer to this question due to the lack of dynamical detail. We can only understand the change mathematically.

The distribution, Eq. 25, includes three factors, $Y^{\alpha\gamma}$, $\frac{1}{2\sqrt{\pi Dt}} \exp(-\frac{[Y^{-\alpha(\gamma-1)t - \phi t + \gamma DY^{-\alpha(\gamma-1)t}]^2}{4Dt})$ and $\exp(-\phi\gamma Y^{-\alpha(\gamma-1)t} + D\gamma(2\gamma-1)Y^{-\alpha(2\gamma-2)t})$. The first factor indicates a growth in power-law. The second factor is the leading factor and actually the Gaussian distribution, which is modulated by the first and the third ones. The second and the third are similar to those corresponding factors in Eq. 22. However, the first varies in the direction opposite to that of Eq. 22, decaying in power-law. This can mathematically interpret the cause why the GDP distribution varies in the direction (from the SPL to the Gaussian in the supercritical situation) of which is completely opposite to that of the population distribution (from the Gaussian to the SPL in the subcritical situation).

Discussion

In this paper, the empirical investigation on the regional populations and GDPs is carried out, and reveals that the distribution of the regional populations and that of the regional GDPs obey the similar regularities, the SPL.

To understand these distributions theoretically, we suggest a generalized Langevin equation to describe the evolution of regional population driven by the deterministic economic factors such as fixed capital investment, flowing capital investment, marginal propensity to save, potential capacity, as well as their random fluctuations. The generalized Langevin equation is transformed into the F-P equation, by a mathematical technique, to describe the evolution of the population distribution. The solution shows a gradual transformation of the distribution from the normal Gaussian distribution at the evolutionary beginning to the SPL one as it is beyond a fixed time $t = 13$. The fixed time is actually a critical point of time. As it is supercritical, the SPL sets up and then remains qualitatively unchanged. Obviously, this situation corresponds to the current practical observations.

Relating the regional GDP to the regional population by introducing a relation function based on the well-known C-D production function, we obtain the distribution of the regional GDPs. The distribution demonstrates the variation, in the supercritical situation, from the SPL to the Gaussian.

As far as the meanings of these distributions are concerned, the variation of the population distribution implies that the

world evolves from a disorder or an unstructured "primitive society" marked by the Gaussian distribution to an order and highly inhomogeneous structured "modern society" characterized by the SPL distribution. How can this SPL distribution set up? The answer may lie in the development pattern of human society, attracting the population by economy/resources and developing the economy/enriching the resources by population reversely. Therefore, the regions of rich resources become the population gathered regions and the social structure expressed by the SPL distribution of populations gradually sets up. The GDP distribution indicates a transformation in the supercritical situation, from the SPL to the Gaussian. This implies that our society is currently in the time point near to the critical point due to the two distributions of populations and GDPs are all the SPL and qualitatively in accordance with the observations. And then one can predict that our society might evolve into the one, in future, dominated by the Gaussian distribution of regional GDPs. Furthermore, the evolution of the world might tend to becoming "common prosperity" denoted by the distribution peak moving to the right, the peak value dropping and the range broadening. Perhaps, someone raises the question of whether there is a contradiction between the two distributions in the supercritical situation—one is the SPL, the other is the Gaussian. The answer may be no due to the basic fact that the growth mode of the GDP is gradually changes from the previous resource-based to the future science-technology-based. With scientific and technological advancements, the close correlation between the production and population becomes invalid; the regional difference in the economy scale reduces or eliminates. Therefore, the relative uniform distribution of the regional GDPs denoted by the Gaussian distribution will occur naturally. Of course, this change might not change the distribution of the regional populations, maintaining the SPL.

Merging the evolution of the population distribution with that of the GDP distribution, a historical picture of human society spreads in front of us. This might have important academic values for enhancing the knowledge about the evolution of our society, population and the related economy.

Acknowledgements

This study is supported by National Natural Science Foundation of China (Grant Nos. 11265011 and 11665018).

Author contributions statement

All authors participated in the design of the project. J. H. and X. -M. W. wrote the manuscript. J. H. and P. W. collected data and analysis. J. H., X. -M. W. and R. H. discussed the model. J. H. and X. -M. W. did theoretical analysis. X. -M. W. supervised the research.

References

1. Ravenstein, E. G. The laws of migration. *Journal of the Statistical Society of London* **48**, 167-235 (1885).
2. Heberle, R. The causes of rural-urban migration a survey of german theories. *American Journal of Sociology* **43**, 932-950 (1938).
3. Bogue, D. J. *Internal migration* (1959).
4. Lee, E. S. A theory of migration. *Demography* **3**, 47-57 (1966).
5. Schultz, T. W. Investment in human capital. *The American economic review* **51**, 1-17 (1961).
6. Zipf, G. K. The $P \propto 1/P^2$ hypothesis: on the intercity movement of persons. *American sociological review* **11**, 677-686 (1946).
7. Stouffer, S. A. Intervening opportunities: a theory relating mobility and distance. *American sociological review* **5**, 845-867 (1940).
8. Domencich, T. A. & McFadden, D. Urban Travel Demand: A Behavioral Analysis. *Canadian Journal of Economics/revue Canadienne D'economique* (1975).
9. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462-465 (2006).
10. Gonzalez, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779-782 (2008).
11. Xiao-Pu, H., Qiang, H., Bing-Hong, W. & Zhou, T. Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **83**, 036117 (2011).
12. Candia, J. *et al.* Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* **41**, 224015 (2008).

13. Wei, H., Xiao-Pu, H., Tao, Z. & Bing-Hong, W. Heavy-tailed statistics in short-message communication. *Chinese Physics Letters* **26**, 028902 (2009).
14. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nature Physics* **6**, 818-823 (2010).
15. Wu, Y., Zhou, C., Xiao, J., Kurths, J. & Schellnhuber, H. J. Evidence for a bimodal distribution in human communication. *Proceedings of the national academy of sciences* **107**, 18803-18808 (2010).
16. Zhi-Dan, Z., Hu, X., Ming-Sheng, S. & Tao, Z. Empirical analysis on the human dynamics of a large-scale short message communication system. *Chinese Physics Letters* **28**, 068901 (2011).
17. Rambaldi, S., Bazzani, A., Giorgini, B. & Giovannini, L. Mobility in modern cities: looking for physical laws. In *Proceedings of the ECCS*, vol. **7**, 132 (ECCS, 2007).
18. Jiang, B., Yin, J. & Zhao, S. Characterizing the human mobility pattern in a large street network. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **80**, 021136 (2009).
19. Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R. & Giovannini, L. Statistical laws in urban mobility from microscopic GPS data in the area of florence. *Journal of Statistical Mechanics: Theory and Experiment* **2010**, P05001 (2010).
20. Jiang, B. & Jia, T. Exploring human mobility patterns based on location information of us flights. *arXiv preprint arXiv:1104.4578* (2011).
21. Liang, X., Zheng, X., Lv, W., Zhu, T. & Xu, K. The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics & its Applications* **391**, 2135-2144 (2012).
22. Peng, C., Jin, X., Wong, K.-C., Shi, M. & Liò, P. Collective human mobility pattern from taxi trips in urban area. *PloS one* **7**, e34487 (2012).
23. Riccardo, G., Armando, B. & Sandro, R. Towards a statistical physics of human mobility. *International Journal of Modern Physics C* **23**, 1250061 (2012).
24. Xiao-Yong, Y., Xiao-Pu, H., Tao, Z. & Bing-Hong, W. Exact solution of the gyration radius of an individual's trajectory for a simplified human regular mobility model. *Chinese Physics Letters* **28**, 120506 (2011).
25. Xiao-Yong, Y., Xiao-Pu, H., Bing-Hong, W. & Zhou, T. Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific reports* **3** (2013).
26. Jie, H., Xu-Ming, W., Ning, Z. & Rui, H. Statistical characteristics of dynamics for population migration driven by the economic interests. *Physica A Statistical Mechanics & Its Applications* **451**, 123-134 (2016).
27. Tao, Z., Xiao-Pu, H., Xiao-Yong, Y. *et al.* Statistical mechanics on temporal and spatial activities of human. *Journal of University of Electronic Science & Technology of China* **42**, 481-540 (2013).
28. Hu, Y., Zhang, J., Huan, D. & Di, Z. Toward a general understanding of the scaling laws in human and animal mobility. *EPL (Europhysics Letters)* **96**, 38006 (2011).
29. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96-100 (2012).
30. Xu-Ming, W., Peng, W., Ping, Z., Rui, H. & Jie, H. Statistical dynamics of early river networks. *Physica A: Statistical Mechanics & Its Applications* **391**, 4497-4505 (2012).
31. Rinaldo, A., Rodriguez-Iturbe, I. & Rigon, R. Channel networks. *Annual review of earth and planetary sciences* **26**, 289-327 (1998).
32. Chechkin, A. V. & Klages, R. Fluctuation relations for anomalous dynamics. *Journal of Statistical Mechanics: Theory & Experiment* **2009**, L03002 (2009).
33. Chavanis, P.-H. Nonlinear mean field fokker-planck equations. application to the chemotaxis of biological populations. *The European Physical Journal B* **62**, 179-208 (2008).
34. Alber, M., Chen, N., Lushnikov, P. M. & Newman, S. A. Continuous macroscopic limit of a discrete stochastic model for interaction of living cells. *Physical review letters* **99**, 168102 (2007).
35. Cobb, C. W. & Douglas, P. H. A theory of production. *The American Economic Review* **8**, 139-165 (1928).
36. Domar, E. D. Capital expansion, rate of growth, and employment. *Econometrica* **14**, 137-147 (1945).
37. Domar, E. D. Expansion and employment. *The American Economic Review* **37**, 34-55 (1947).

38. Barabási, A., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. & Vicsek, T. Evolution of the social network of scientific collaborations. *Physica A Statistical Mechanics & Its Applications* **311**, 590-614 (2002).
39. Chang, H., Bei-Bei, S., Yue-Ping, Z. & Da-Ren He. Assortativity and act degree distribution of some collaboration networks. *Physica A: Statistical Mechanics & Its Applications* **383**, 687-702 (2007).
40. Raup, D. M. Biological extinction in earth history. *Science* **231**, 1528-1533 (1986).
41. Laherrère, J. & Sornette, D. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *Physics of Condensed Matter* **2**, 525-539 (1998).